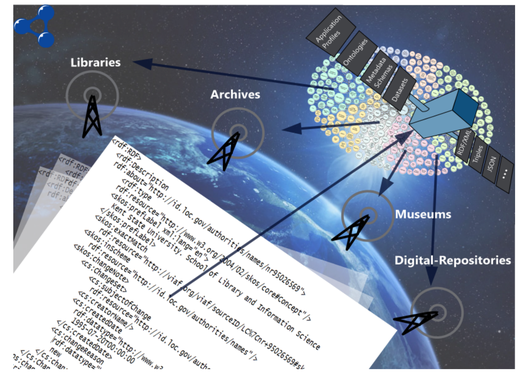
Archival description and linked data: Opportunities and implementation challenges

Karen F. Gracy, Ph.D., Kent State University



The Metadata Vocabulary Junction Project

http://lod-lam.slis.kent.edu

This study aims to show how libraries, archives, and museums (LAMs) can benefit from the data and metadata resources that have been created as a result of the Linked Open Data (LOD) movement.

Background

The LOD movement seeks to connect related data formerly isolated in repositories (often called "silos") and not previously linked. Cultural heritage institutions generate valuable data through tools such as finding aids, cataloging records, and databases, but this information is often sequestered inside archival information systems that cannot be easily linked to other data sources.

Linked data technologies and LOD resources can enable LAMs to …

- ✓ Enhance their already existing cataloging records, finding aids and other services;
- ✓ Help users obtain new information and knowledge relevant to materials in their institution; and,
- ✓ Make their collections available to users of external databases.

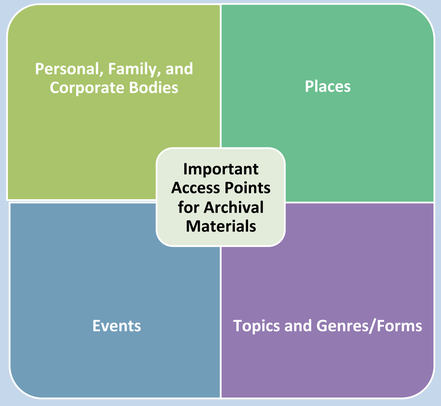
Aligning archive data with LOD data constructs

Archive data sources:

- Bibliographic records for archival collections
  - MARC records, schema.org markup
- Archival description
  - EAD-encoded finding aids

LOD data sources:

- Ontologies
  - e.g., GeoNames, EAC-CPF
- Mark-up/metadata elements sets
  - e.g., schema.org, FOAF, LODE
- Structured data in XML/RDF samples
  - e.g., Dbpedia
- Associated documentation



Data collection and analysis methods

Archival information systems

1. **Analyze descriptive and authority standards to determine what information may be useful as linked data.**
2. **Review record exemplars from several sources to discover how standards are applied.**
3. **Identify and review other potential sources of information as needed.**
4. **Break down the records and find:**
   - **Common elements**
   - **Possibilities for linking archival data elements to LOD classes and properties**
   - **Hidden access points**

LOD datasets

1. **Identify relevant datasets**
2. **Analyze the source of the data structures**
   - **Ontologies used**
   - **Metadata schemas and application profiles**
   - **Structured data in XML/ RDF examples**
   - **Documentation**
3. **Crosswalk their matching properties, indicating the matching level (broader, narrower, equivalent, or close match)**
4. **Identify major classes and properties relevant to archival data**

Partial alignment example
EAD to schema.org

Root: EAD tag element	Target: schema.org Class or property	Property Mapping
<corpname>	Organization	N
<geogname>	Place	CM
<genreform>	genre	N
<scopecontent>	description	B

Significant Findings

1. Archival encoding standards tend to align more closely at the class level for Dbpedia, GeoNames, and schema.org, (i.e., the latter are more granular than archival description).

2. FOAF and LODE are more likely to match archival description at the property level (thus have similar levels of granularity to archival standards).

3. There is a significant need for increasing the specificity of EAD tags through use of attributes and adding tags that would allow greater precision.

Research funded through the IMLS National Leadership Grant program.

INSTITUTE of Museum and Library SERVICES

Principal Investigator: Marcia Lei Zeng
Co-Investigator: Karen F. Gracy